DELAD-CLARIN Workshop 22-23 September 2022: Introduction

Henk van den Heuvel, CLST Radboud University

CLST | Centre for Language and Speech Technology
Radboud University

Practical matters

- *Please note that **this event will be recorded**.*

- ***If you do not wish to be recorded**, please make sure that during the event your microphone is muted, your camera is turned off and you ask questions by using the chat function only.*

- *Edited versions will be made available for CLARIN dissemination*
  - Presentations will be collected in Googlefolder (Maria?)

  - please mail to Henk.vandenheuvel@ru.nl

- *Group Photo before coffee break*

DELAD initiative

- Initiative to collect and share corpora of speech with disorders (CSD):
- http://delad.net/

- Partners are a mix of researchers, infrastructure specialists, legal experts

- DELAD organises annual workshops since 2015 where these groups convene

- Since 2017 under CLARIN header and support

Topics addressed:
- Examples of CSD
- Guidelines for collecting and sharing CSD
- Ethics and legal aspects
- Levels of anonymisation
- Layered access of data
- Integration of CSD in the CLARIN infrastructure
- Formats
- Relevant metadata

DELAD Background

- Second CLARIN-DELAD Workshop, Utrecht NL, SURF premises, 28-29 Jan. 2019
  - Reaffirmation that CLARIN is the Data Trust, to provide the data fence around CSD.
  - DELAD should apply to become a Task Force within CLARIN focusing on practical issues on sharing CSD.
  - As a result, the DELAD website should be updated with a CLARIN flag and contain relevant guidelines for collecting, sharing and storing CSD.
  - Talkbank is seen as a good CLARIN site to host CSD, especially if a European storage cloud and stricter access policy can be realised.
  - Collaboration DELAD, ACE, TLA, Talkbank in data curation

- Third CLARIN-DELAD Workshop, Online 27-28 Jan. 2021
  - Extend DELAD network with new participants
  - Explore with the participants the potential of the new CLARIN K-Centre on Atypical Communication Expertise (henceforth ACE) for hosting CSD of DELAD members
  - Exchange deeper insights on Data Protection Impact Assessments (DPIAs),
  - Discuss voice conversion as a means to pseudonymize speech

DELAD CLARIN Proposal

Goals of this workshop:

- Extend DELAD network with new participants

- Research & CSD presentations

- Obtaining clinical data via hospitals vs Obtaining data outside clinical institutes from alternative

- The impact of the Data Governance Act & Data Altruism

- Continued discussion on voice conversion as a means to pseudonomise speech.

Programme for 22 September

- 13:00 – 13:15 Introduction (Henk van den Heuvel)

- 13:15 – 14:25 Research presentations (Chair Satu Saalasti):
  - "Speech STAR - An outline ultrasound- tongue-imaging-based curated corpus of disordered and non-disordered speech" (Eleanor Lawson, Joanne Cleland, & Jane Stuart-Smith)

  - "Using video-based pose estimation for automated analysis of interaction" (Satu Saalasti)

  - "Rethinking Language: Computational approaches to speech and language assessment for clinical studies" (Chiara Barattieri di San Pietro)

  - Group discussion
- 14:25 – 14:40 Coffee break / group picture

Programme for 22 September

- 14:40 – 15:40 Recent developments at ACE / DELAD for support of sharing CDS (DELAD Steering group)

- 15:40 – 16:00 Follow up on voice Pseudonymisation (Rob van Son) (pre-recorded presentation) (Chair: Henk van den Heuvel)

- 16:00-16:20 "Crosslinguistic data in typical and protracted phonological development" (Barbara Bernhardt)

Programme for 23 September

- 13:00 – 13:15 Introduction (Henk van den Heuvel)

- 13:15 – 14:25 Research presentations (Chair Nicola Bessell):
  - "Developments in sensitive data processing at CSC" (Martin Matthiesen)

  - "The assumptions for the analysis of speech disorders and primary functions using the CARSTENS AG501 articulograph and an acoustic field distribution analyzer" (Katarzyna Klessa, Anita Lorenc, Łukasz Mik, Agnieszka Borowiec, & Daniel Król)

  - "Licensing and distributing language resources via the Language Bank of Finland" (Mietta Lennes)

  - Group discussion
- 14:25 – 14:40 Coffee break / group picture

Programme for 23 September

- 14:40 – 15:40 The impact of the Data Governance Act & Data Altruism (Pawel Kamocki) (Chair: Alice Lee)

- 15:40 – 16:40 Lessons learnt from the use of the data sharing template of Dutch Academic hospitals (Anne Jan Sikkema) (Chair: Kataryna Klessa)

- 16:40- A chat over drinks